Article

Correlational analysis between feed amino acid profile and heamatological parameters of Atlantic salmon (*Salmo salar*)
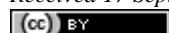
Familusi Oluwatosin Adekunle[1], **I. A. Adebayo**[1], **Obaisi Alaanuloluwa**[2]

[1]Fisheries and Aquaculture Management Department, Ekiti-State University, Ado-Ekiti, Nigeria
[2]Environmental Consultant, Thomas Chase Africa, Ado-Ekiti, Nigeria
E-mail: familusiadekunle@gmail.com, Isreal.adebayo@eksu.edu.ng, obaisiai@gmail.com

Abstract

Sustainable use of natural resource such as fishmeal in aquaculture, can be achieved by substituting with more abundant biomass. This goal relies on the use of available research methods to understand how cultured species respond to each substitute and factors responsible for such responses to know how they can be modified for better response. The quantity and quality of research needed for understanding these relationships can be optimized by employing identified statistical methods. Correlation coefficient is a statistical method to quantify degree of corresponding changes between 2 features, while LASSOCV gives a defined list of independent features, relevant to the target feature being analyzed based on the ordinary least squares. Most relevant features to RBC are phenylanaline and valine, other less relevant features are histidine and threonine, all other essential amino acids are identified as irrelevant to RBC with a certainty score of 50% (R^2). RBC and HB showed strong relationship with a coefficient of 0.8, but differ in relevant amino acid components (isoleucine, histidine and threonine). The analysis suggests the relevant amino acids for each clinical parameters, which can be confirmed with further experiments.

Keywords correlation coefficient; LASSOCV; RBC; HB; amino acid.

1 Introduction

Increasing intensification in aquaculture demands continuous need for feed suitability assessment using clinical parameters such as heamatology by the re-evaluation of nutrient recommendations. The aquaculture industry depends heavily on wild catches, hence sustainable marine resource use can only be achieved by the inclusion of non-catch dependent biomass as the source of animal protein in salmon production. Ongoing research focuses on the inclusion of different levels of a particular protein substitute alongside other highly controlled factors to arrive at fish-meal substitute recommendations, as done by Hevroy et al. (2005), Hemre et al. (2016), Alex et al. (2016), Sagstad et al. (2008), and Metochis et al. (2016).

1.1 Correlational analysis

Correlational analysis is a measure of the extent to which two variables change together. It also shows the strength and direction of the relationship (negative or positive) (Zhang, 2018; Zhang and Li, 2015). It is the simplest statistical technique for analyzing causal effects. The correlation coefficient measures the linear relationship and is also referred to as Pearson product-moment correlation coefficient (Marsden, no-date).

Regression problems are quite straightforward and the selection of predictor variables that best predict the outcome of particular features can be challenging. As a result, Bradely Efron identified predictor variable selection in statistics. There's a lot of statistical literature on refining statistical methods to better select predictor variables, however, the most widely researched and implemented modern methods is the least absolute shrinkage and selection operator (LASSO). This fits within the broader least angle regression framework (LARS) that can estimate LASSO with the computational complexity of ordinary least squares (OLS) (Daniel, M.2015).

OLS is used to predict a metric outcome from many predictors, in this study, the predictors comprise of amino acids included in the feed of Atlantic salmon. Coefficient estimated are obtained by minimizing it's function so that the vertical squared distances between the observed data and predicted values are smallest. Mathematically, loss function can be represented as:

$$L^{OLS}(\beta) = Y - X\beta^2$$

where,

> L = Loss function
> X = n × p design matrix of predictors
> Y = n × 1 vector of responses
> β = p × 1 vector of regression coefficients.

If the feature is irrelevant, LASSO penalizes it's coefficient and makes it 0. Hence the features with coefficient '0' are removed and the rest are taken. According to Wilkinson (1979), in feature selection procedures, overfitting is magnified and both T an F statistics do not follow the appropriate distributions because standard inferential tests and their associated p-values are not appropriate for repeatedly swapping predictors in and out of the model while significance testing at the same time. This can result in Type 1 error rates for hypothesis testing on regression coefficients being too large, resulting in the inclusion of more features in the model.

LASSO utilizes penalty that is a function of the regression coefficient by setting an absolute value of the sum of coefficients which will shrink the coefficient. The LASSO penalty is represented as a penalized OLS loss function as described by Lockhart et al. (2014)

$$L^{Lasso}(\beta) = Y - X\beta^2 + \lambda W^T \beta$$

where

L = loss function
X = n × p design matrix for predictors
Y = n ×1 vector of responses
β = p ×1 vector of regression coefficients
W = p ×1 vector of ±1 values that are congruent with the sign of the associated element of β
$\lambda \geq 0$ = regularization parameter that controls the degree of shrinkage

1.2 Amino acid profile

Fishmeal is an essential component of aquatic feeds, however, competition for fishmeal by demand from the terrestrial livestock industry and human consumption has led to seeking alternatives for fish feed. Alternatives are sought from grains, oilseeds, animal by-products, industrial wastes and unicellular organisms (yeasts, mold and bacteria) (Christoforos et al., 2016). Suitability of any of the substitute depends on several factors including anti-nutritional factor composition, amino acid profile, digestibility among several other factors. Research based on purified diets showed that Atlantic salmon required 45 percent protein (Lall and Bishop, 1977).

Atlantic salmon is the most successfully farmed salmonids but the nutrient requirement are not well defined (Santosh, 2011). In fish, the quality of protein alternatives depends on amino-acid composition, digestibility and availability. To some degree, amino-acid quality can be considered as the extent of similarity between the alternative and the balance of amino-acid required by the animal (the IAA) (Wang and Fuller, 1989). IAA is assessed by dose-response experiments (Wilson, 2003), which is not only time and resource intensive, but also prone to methodical problems and the results reported may be inconsistent with other experiments. As reported by Xavier et al (2003) and Santosh (2011), IAA of Atlantic salmon include arginine, histidine, isoleucine, leucine, lysine, methionine/cysteine, phenylalanine/tyrosine, threonine and valine (Abbreviation: IAA: Idespensable Amino Acid, AA: Amino Acid).

1.3 Haematology of Atlantic salmon

Blood component is a useful indicator of the physiological status and health of salmon (Alex et al., 2015). It plays a key role in aquaculture because of its importance in monitoring fish health condition. The parameters are important in establishing a standard value range, deviation from which may confirm a disturbance in the physiological process. These haematological parameters are affected by factors including age, species, environment, response of condition, age, maturation, diet and nutrition (Nasser and Yasser, 2018)

Heamoglobin (HB) is made up of protein in red blood cells responsible for delivery of oxygen to the tissues. To ensure continuous supply of oxygenated blood, an optimal level of HB must be maintained. The RBC are nucleated, oval shaped biconvex disc, loaded with haemoglobin and functions primarily to transport oxygen as it contains the pigment haemoglobin. Haematocrit (HCT) is also referred to as the PCV, and is a part of the complete blood count (CBC) which denoted the percent of the blood made up of red blood cells. Mean Corpuscular Volume (MCV) is the average RBC volume. It is a calculated value derived from HCT and RBC. Mean Corpuscular Haemoglobin (MCH) is the average amount of HB in the RBC. MCHC or mean corpuscular haemoglobin concentration is the average concentration of HB in a given volume of RBC (Surjya et al., 2018)

2 Materials and Methods

2.1 Data acquisition

Data on amino acid profile ofmarine protein, marine oil and marine-free diets reported by Metochis et al. (2016), the composition of fish protein hydrolysate reported by Hevrøy et al. (2005), and the AA composition of *Palmaria palmata* as reported by Alex et al. (2016) are collated alongside respective clinical response figures. Data is entered into a single excel file, loaded into a jupyter notebook on the anaconda package navigator and analyzed using pandas, seaborn and matplotlib's package's documentation.

2.2 Amino acid estimation

Analysis is on arginine, histidine, isoleucine, leucine, lysine, methionine/cysteine, phenylalanine/tyrosine, threonine and valine. Selected parameters are the Indispensable Amino Acids). Amino Acid laboratory estimation for data used in the study followed the procedures aligned by Cohen et al. (1989).

2.3 Clinical parameters

Haematocrit(%), haemoglobin, red blood cell count (10^{12} L^{-1}), mean cell volume (10^{-15} L), haemoglobin (g 100 mL^{-1}), mean cell concentration (10^{-6} L), mean cell haemoglobin concentration (g 100 mL^{-1}). Clinical parameter estimated are similarly outlined in selected research following the procedure outlined by Sandnes et al. (1988).

2.4 Preliminary data

Preliminary date are listed in Tables 1 and 2.

Table 1 Amino acid profile of experimental diets.

Arg	His	Iso	Leu	Lys	Met	Phe	Thr	Val
0.89	0.81	0.82	0.87	0.91	0.83	0.84	0.81	0.83
0.89	0.8	0.81	0.86	0.91	0.83	0.84	0.79	0.81
0.9	0.82	0.82	0.87	0.92	0.84	0.85	0.8	0.82
0.91	0.83	0.84	0.89	0.92	0.85	0.86	0.83	0.85
0.92	0.84	0.86	0.89	0.93	0.87	0.87	0.84	0.86
0.92	0.83	0.85	0.89	0.93	0.87	0.87	0.84	0.85
0.7	0.195	0.37	0.79	0.74	0.25	0.37	0.43	0.47
0.73	0.19	0.41	0.79	0.75	0.23	0.41	0.45	0.48
0.72	0.22	0.36	0.8	0.73	0.23	0.38	0.45	0.45
0.71	0.17	0.37	0.78	0.73	0.23	0.41	0.43	0.44
0.76	0.21	0.41	0.81	0.74	0.22	0.41	0.46	0.47

Arg-   Arginine, His – Histidine, Iso – Isoleucine, Leu – Leucine, Lys – Lysine, Met – Methionine,   Phe – Phenylalanine,

Thr – Threonine, Val - Valine

Table 2 Corresponding haematological responses.

RBC	HCT	HB	MCV	MCH	MCHC
0.11		0.37	0.75	0.3252	0.661	0.203
0.11		0.37	0.73	0.3251	0.639	0.196
0.12		0.383	0.8	0.318	0.666	0.21
0.11		0.372	0.75	0.324	0.651	0.201
0.11		0.374	0.73	0.3282	0.643	0.196
0.11		0.347	0.7	0.316	0.643	0.204
0.106		0.378	0.72	0.358	0.68	0.19
0.11		0.379	0.75	0.345	0.68	0.199
0.11		0.383	0.75	0.35	0.69	0.197
0.114		0.368	0.8	0.326	0.7	0.217
0.11		0.366	0.74	0.335	0.67	0.201

RBC –Red Blood Cell, HCT- Haematocrit, HB-Heamoglobin, MCV- Mean Corpuscular Volume, MCH-Mean

Corpuscular Haemoglobin , MCHC- Mean Corpuscular Haemoglobin Concentration

2.5 Data analysis

Using historical dataset on feeding trials conducted on *Salmo salar*, data is loaded into excel workbook then jupyter using the read_excel method. Independent features are selected using the filter method. The filter method is based on selecting a subset of the independent variables depending on how related it is to each of the haematological feature being estimated. The filtering is done using a correlation matrix based on Pearson's. The correlation heatmap is plotted and annotated following the sea born package's documentation. Annotation on the heatmap (correlation coefficient) indicates the degree of relatedness of two intersecting variables. The coefficient has a maximum value of 1 and a minimum value of -1

- Value close to 1 implies the stronger positive correlation or relatedness and an increase in the independent feature follows an increase in the dependent or target feature.
- Value close to -1 shows stronger negative relatedness and an increase in the independent feature is accompanied by a decrease in the dependent or target feature.
- Value close to 0 shows unrelatedness between the two features

Correlation coefficient with values >0.5 according to the heatmap are selected for each heamatological feature while other features with coefficient >0.5 are dropped. The assumption of linear regression is that independent variables need to be uncorrelated with each other. If these variables are correlated, it is most appropriate to keep one for further analysis. Therefore, following estimation of the most relevant features for each haematological parameter, the resulting independent features is checked for relatedness using the same correlation matrix on the heatmap.

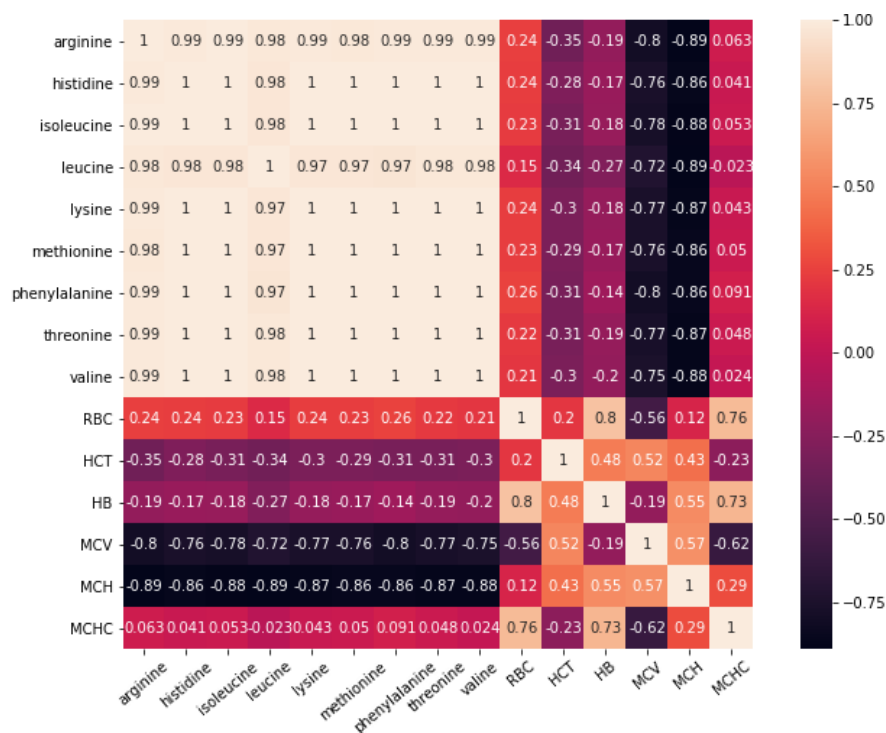**3 Results**

The results are shown in Figs. 1-7.

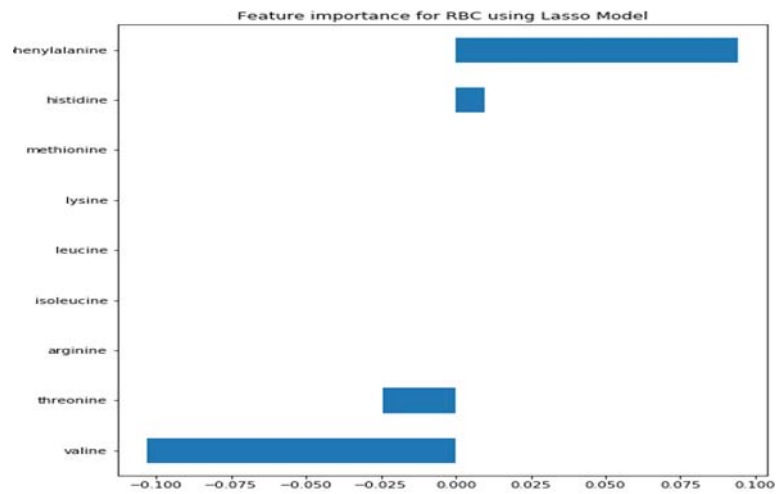**Fig. 1** Annotated correlation heatmap.

Fig. 2 RBC LASSOCV.

In Fig. 2, best alpha using built-in LassoCV: 0.000005; best score using built-in LassoCV: 0.500719, and Lasso picked 4 variables and eliminated the other 5 variables.
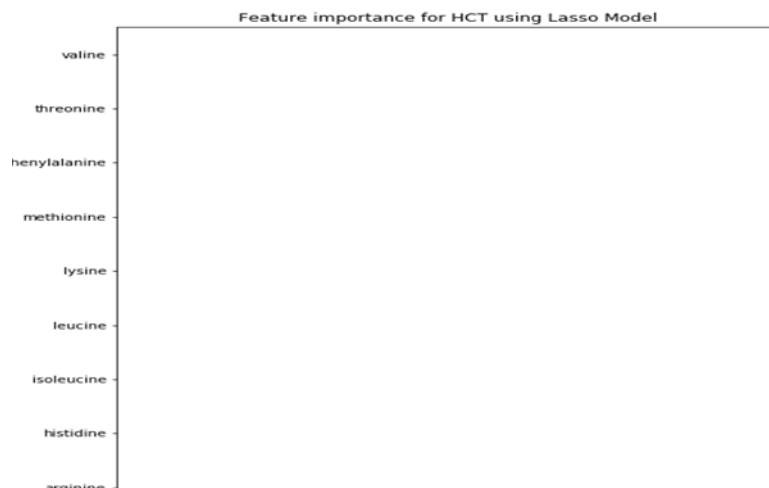

Fig. 3 HCT LASSOCV.

In Fig. 3, best alpha using built-in LassoCV: 0.000869; best score using built-in LassoCV: 0.0000 00, and Lasso picked 0 variables and eliminated the other 9 variables.
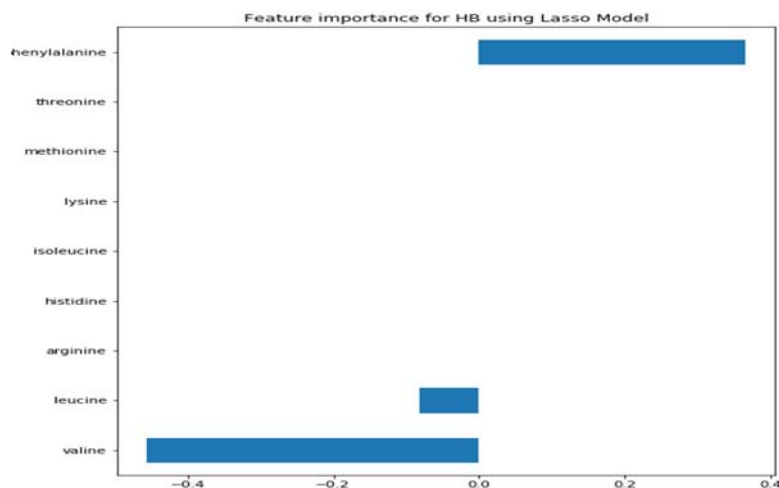
Fig. 4 HB LASSOCV.

In Fig. 4, best alpha using built-in LassoCV: 0.000093; best score using built-in LassoCV: 0.294310, and Lasso picked 3 variables and eliminated the other 6 variables.
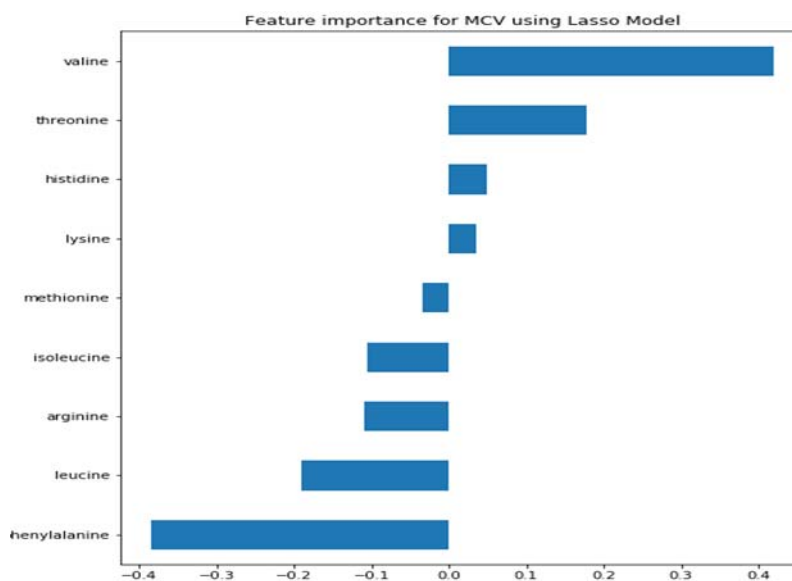

Fig. 5 MCV LASSOCV.

In Fig. 5, best alpha using built-in LassoCV: 0.000003; best score using built-in LassoCV: 0.940455, andLasso picked 9 variables and eliminated the other 0 variables.
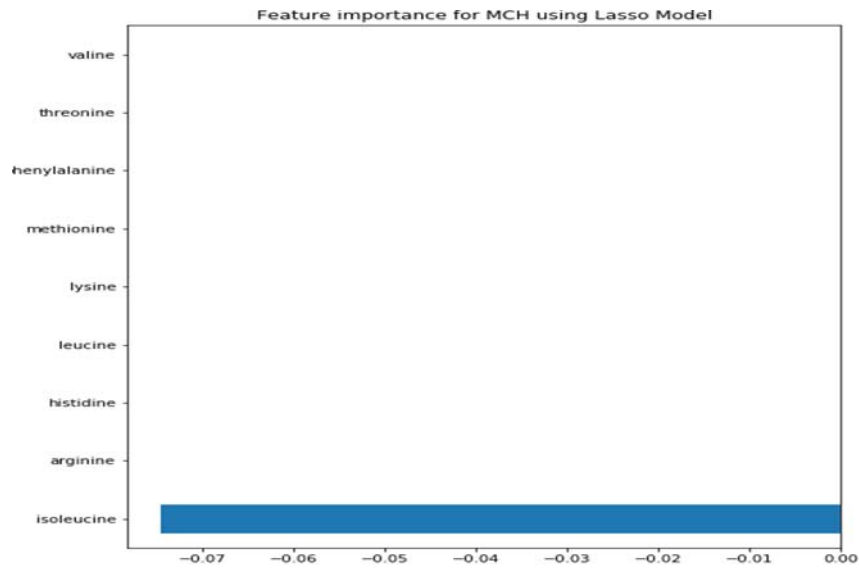
Fig. 6 MCH LASSOCV.

In Fig. 6, best alpha using built-in LassoCV: 0.000080; best score using built-in LassoCV: 0.769 351, and Lasso picked 1 variables and eliminated the other 8 variables.
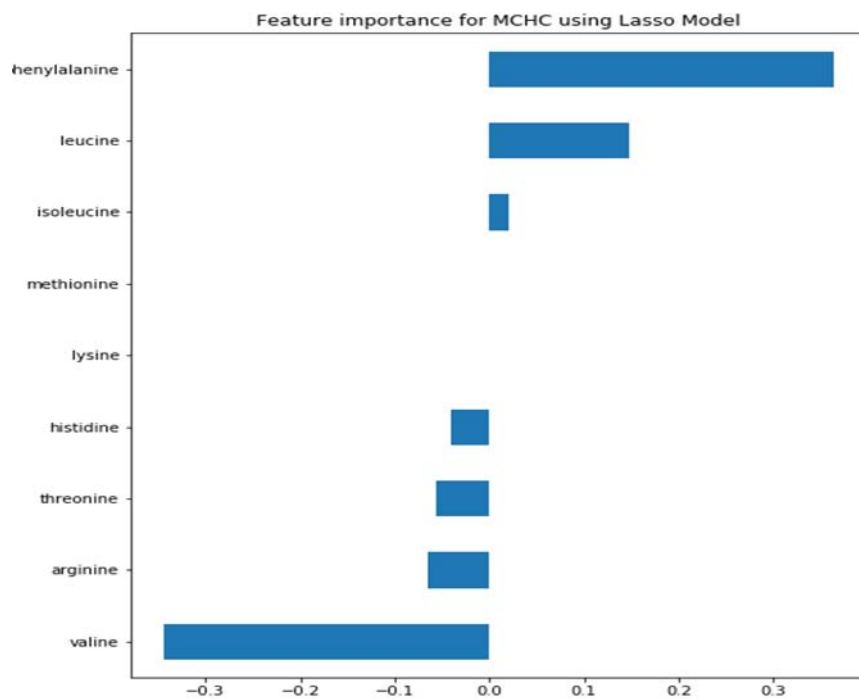

Fig. 7 MCHC LASSOCV.

In Fig. 7, best alpha using built-in LassoCV: 0.000004; best score using built-in LassoCV: 0.710457, and Lasso picked 7 variables and eliminated the other 2 variables.

4 Discussion

RBC feature showed the highest relationship with phenylalanine, histidine and a negative relationship with threonine and valine. With an accuracy of 50% according to the best LASSO score, the prediction may not be as accurate as possible hence, further experimentation is needed to verify the relationship between the red blood cell and the correlated amino acid. All selected features also show the highest level of correlation above other features on the heatmap plot with coefficient values of 0.26, 0.24, 0.22 and 0.21 respectively. While arginine showed a correlation coefficient similar to histidine. It is expected that cysteine and tyrosine for instance, can be synthesized endogenously from methionine and phenylalanine respectively, both of which are IAA and need to be supplied through the diet according to Syne et al. (2015), therefore, a deficiency of phenylalanine or methionine can result in deficiency symptoms, leading to heamatological response of the RBC.

None of the amino acids were selected as important to HCT. With a score of 0%, the feature showed weak negative correlation with all amino-acids according to the heatmap. Prediction of this variable can be determined with an accuracy of 52% using MCV which is also a target variable. However, arginine resulted in increased HCT, HB and RBC count after mitotic exposure as reported by Pohlenz (2012).

Phenylalanine, leucine and valine are shown to be the most relevant features to HB. With a score of 30, the selected variables should be subject to further testing and experimentation. This is also shown by the weak negative correlation coefficient as shown on the heatmap plot. However, the coefficient of 80% with RBC, phenylalanine and valine are both suspected to be more accurate relevant features to HB since it is relevant to both RBC and HB which strongly correlates. Surjya et al. (2018) noted changes in RBC is accompanied by similar response in HB levels. Conversely, concentrations of RBC and HB are shown not to be affected by varying levels of dietary phenyalanine levels in *Orechromis niloticus* and *O. aureus* (Wei et al., 2018).

With a maximum coefficient of -0.72 and a minimum of -0.8, all the indispensable amino acids are selected as important features to MCV. A high score of 94% also supports the genuity of the feature importance. Of all the variables, MCV can also serve as reference variable for RBC and MCHC with accuracy of 56 and 62% respectively.

The only feature relevant to MCH is isoleucince according to the LASSO feature selection. With a correlation coefficient of 0.54; the highest compared to all other coefficients. Other target variables such as MCV and HB also show positive relationship with MCH, with coefficients of 55 and 57% respectively. An existing dynamic is observed in all the related target variables with the indication of isoleucine as relevant features in all the three target features. The exclusion of methionine and lysine as relevant features to MCHC with an accuracy of 71%can be attributed to the proximity of their correlation coefficients with that of other features like histidine and isoleucine. Correlation of RBC and HB with 76 and 73%, relevant target features shared in common include phenylalanine and valine.

References

Alex H, Soler-Vila A, O'Keeffe D, Casburn P, Richard F, Johnson M. 2015. The inclusion of *Palmaria palmata* macroalgae in Atlantic salmon (*Salmo salar*) diets: Effects on growth, haematology, immunity and liver function. Journal of Applied Phycology, 28(5): 3091-3100

Cohen S, Meys M, Tarvin T. 1989. The Pico Tag Method. A Manual of Advanced Techniques For Amino Acid Analysis. Water Division of Millipore WMO2, Rev 1 p 123, Milford, MA, USA

Hevrøy E, Espe M, Waagbø R, Sandnes K, Ruud M, Hemre G. 2005. Nutrient utilization in Atlantic salmon (*Salmo salar*) fed increased levels of fish protein hydrolysate during a period of fast growth. Aquaculture Nutrition, 11: 301-313

Hemre G. Erik-Jan L.Pål A. Kristin H. Marit E. Bente E. Torstensen, et al. 2016. Atlantic salmon (*Salmo salar*) require increased dietary levels of B-vitamins when fed diets with high inclusion of plant based ingredients. Peer J, 1-28

Lockhart R, Taylor J, Tibshirani J, Tibshirani R. 2014. A Significance test for the lasso. The Annals of Statistics, 42: 413-468

Lall P, Bishop F. 1977. Studies on mineral and protein utilization by Atlantic salmon (*Salmo salar*) grown in sea water. Fish. Mar. Serv. Tech. Rep, 1-16

Marsden E. Modelling correlations with Python and SciPy (Online publication). http://www.risk-engineering.org. Accessed on 07/25/2020

McNeish D. 2015. Using Lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. Multivariate Behavioral Research, 50(5): 471-484

Metochis C, Crampton V, Ruohonen K, El-Mowafi A, Bell J, Alexandra A, Kimberly D. 2016. Effects of marine protein-, marine oil and marine-free diets on the growth performance and innate immune responses of Atlantic salmon (*Salmo salar*) post-smolts. Aquaculture Research, 1-21

Pohlenz C, Buentello A, Criscitiello M, Smith R, Gatlin D. 2012.Synergies between vaccination and dietary arginine and glutamine supplementation improve the immune response of channel catfish against *Edwardsiella ictaluri*. Fish Shellfish Immunology, 33(3): 543-551

Sandnes K, Lie O, Waagbo.1988. Normal ranges of some blood chemistry parameters in adult farmed Atlantic salmon, *Salmo salar*. Journal of Fish Biology, 32: 129-136

Santosh L. 2011. Atlantic Salmon (*Salmo salar*, Linnaeus 1758). Aquaculture Service (FIRA), FAO Fisheries and Aquaculture Department, FAO, Rome, Italy

Sagstad A, Sandem M, Krogdahl A, Bakke-McKellep M, Fryostad M, Hemre I. 2008. Organs development, gene expression and health of Atlantic salmon (Salmosalar L.) fed genetically modified soybeans compared to the near isogenic non-modified parental line. Journal of Aquaculture Nutrition. 14(1): 556-572

Surjya N, Ajeet S, Amit M, Geeta J. 2018. Effect of different dietary protein sources on hematological parameters of striped catfish *Pangasianodon hypophthalmus*. Journal of Entomology and Zoology Studies, 6(2): 3198-3202

Wang T, Fuller M. 1989. The optimum dietary amino acid pattern for growing pigs: 1. Experiments by amino acid deletion. British Journal of Nutrition. 62(1): 77-89

Wilson R. 2003. Amino acid requirements of finfish and crustaceans. In: Amino Acids in Animal Nutrition 2. 427-447, Wallingford, Oxon, CAB International, UK

Wei X, Zhiying Z,Dayu L, Jinglin Z, Yirong Y, Hong Y. 2018. Effect of dietary phenylalanine level on growth performance, body composition, and biochemical parameters in plasma of juvenile hybrid tilapia, *Oreochromis niloticus Oreochromis aureus*. Journal of the World Aquaculture Society, 1-15

Wilkinson L. 1979. Tests of significance in stepwise regression. Psychological Bulletin, 86: 168-174

Xavier R, Muriel M, Tarik A, Yvan L, Sadasivam J. 2003. The optimum dietary indispensable amino acid pattern for growing Atlantic salmon (*Salmo salar*) fry. British Journal of Nutrition, 90: 865-876

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK

Zhang WJ, Li X. 2015. Linear correlation analysis in finding interactions: Half of predicted interactions are undeterministic and one-third of candidate direct interactions are missed. Selforganizology, 2(3): 39-45